

DETERMINING AN ANALYTIC FUNCTION FROM ITS DISTRIBUTION OF VALUES

PETER WAKSMAN

Let $f(x)$ be a real analytic function defined on a (possibly infinite and possibly closed) interval (A, B) . The frequency distribution of f is defined to be

$$\omega_f(y) = \text{Lebesgue measure } \{x \in (A, B) \mid f(x) \leq y\}.$$

In this paper we consider the problem of determining f given its distribution ω_f . Since a trivial change of f of the form $g(x) = f(a \pm x)$ will have the same distribution, we ask: does ω_f determine f up to such trivial changes? A partial answer is given by

THEOREM. *If f is real analytic with distinct and non-degenerate critical values on a finite interval $[A, B]$ and if the values of f at the endpoints are different from each other, and at least one is different from the value at any critical point on the interior of the interval, then f is determined uniquely (up to trivial changes) by its frequency distribution on this interval. As a consequence we have:*

COROLLARY. *A real analytic function with distinct non-degenerate critical values is determined uniquely (up to trivial changes) by its frequency distribution on the interval between its minimum and maximum critical points.*

The method of proof depends on studying the behavior of the frequency distribution near a critical value, which method fails if the value is repeated at different critical points or if the value is a degenerate critical value ($f(x)$, where $f'(x) = 0$ and $f''(x) = 0$); thus the assumptions which exclude such cases may only be necessary for the proof and not the result. That the assumption of distinct endpoint values is necessary is seen in the following examples.

The functions $f_1(x) = x^2$ on $[-1, 1]$ and $f_2(x) = (x/2)^2$ on $[0, 2]$ both have the frequency distribution $\omega(y) = 2\sqrt{y}$. The critical value 0 is distinct and non-degenerate, but the theorem does not apply to f_1 because $f_1(-1) = f_1(1)$. In general an even function is not determined by its distribution on an interval symmetric about the origin. The function f_2 does satisfy the assumptions of the theorem and is (up to trivial changes) the only such function with distribution $2\sqrt{y}$. It is possible that f_1 is the

only $2-1$ function with this distribution, and it is tempting to try to extend the theorem, replacing the endpoint condition with an assumption about the degree (n , if f is $n - 1$) of the function.

Periodic functions defined on an interval of length equal to a multiple of their period are also not determined by their distributions. For example $\sin(kx)$ on $[0, 2\pi]$ has the same distribution for any integer k . Here the critical values are not distinct (for $k > 1$) and the endpoint values coincide. Again one might hope to extend the theorem by adding an assumption about the degree of the function to be determined; thus it is possible that $\sin(kx)$ is the only analytic function of degree $2k$ with the given distribution. However, as it stands, the theorem only applies when the interval is not of length equal to a multiple of the period, and even then only when the values of critical points in the interior of the interval are distinct and when the endpoint values are different from each other and different from the values at interior critical points. Thus for example $\sin(x)$ is determined uniquely (up to trivial changes) on the interval $[0, 2\pi + \varepsilon]$ where $0 < |\varepsilon| < \pi$.

If an analytic function has repeated critical values but has distinct and non-critical values at the endpoints, the arguments can be modified easily to show the function is determined by its distribution. However, no quick modification will prove the result when the endpoint values are critical values; yet the result might still be true. Thus there is more work to be done on this topic; both for understanding periodic and symmetric cases, as well as for trying to remove the assumption about distinctness and nondegeneracy of the critical values. We do not consider here the case of an infinite interval; in all these cases a version of the theorem may still hold, but other ideas will be needed for its proof.

Motivation. There are general reasons for wanting to view an analytic function as equivalent to its distribution. For example one might want to compute in terms of the distribution rather than the function; or, in an applied context, one might be able to measure the distribution but not the function. The present work was motivated by the following problem in non-linear evolution equations:

If $u(x, t)$ satisfies a differential equation of the form $u_t = F(u)$ (where F can be an operator involving higher derivatives) is the initial condition $u(x, 0)$ determined by the time averages

$$A(t) = \int u(x, t) dx ?$$

In the simplest case, F is a smooth function of u , in which case one can proceed as follows: $A(t)$ determines the derivatives $A^{(n)}(0)$; under mild assumption $A^{(n)}(0) = \int u_{tt\dots t}(x, 0) dx$. By repeated applications of the relation $u_t = F(u)$ this integral can be rewritten as

$$\int \left(F(g) \frac{d}{dg} \right)^{n-1} F(g) dx \quad \text{where } g(x) = u(x, 0).$$

Substituting $y = g(x)$ this becomes

$$\int \left(F(y) \frac{d}{dy} \right)^{n-1} F(y) d\omega_g(y),$$

and in this form the problem is to determine g given the quantities

$$A^{(n)}(0) = \int \psi_n(y) d\omega_g(y),$$

where the $\psi_n(y) = (F(y)(d/dy))^{n-1}F(y)$ depend only on F .

If, for example, the uniform closure of the linear span of $\{\psi_n\}$ contains the continuous functions, then $\omega_g(y)$ is determined at its points of continuity. In this case the initial condition $g(x) = u(x, 0)$ is determined when g is known to satisfy the hypotheses of the theorem.

The paper is organized as follows: In §1 we study the density $d\omega_f/dy$, and prove a preliminary lemma used in the proof of the theorem which comprises §2. The case where one endpoint is not a critical point is considered first and then the case where both endpoints are critical points. Interrupting the proof, there is a short digression about the existence of a function with given distribution.

1. The density $d\omega_f/dy$. In this section we develop an expression for the density $d\omega_f/dy$, which is used later in the proof. Throughout in the following let f be a non-constant analytic function defined on a closed real interval $[A, B]$.

LEMMA 1. *For any measurable function $p(y)$ and non-constant analytic function f defined on $[A, B]$*

$$\int_A^B p(f) dx = \int_{\text{image}(f)} p(y) 1_f(y) dy$$

where $1_f(y) \equiv \sum_{x \in f^{-1}(y)} 1/|f'(x)|$.

REMARK. This is just the change of variables formula when we substitute $y = f(x)$; but that it makes sense when f is not monotonic is called “integration over the fibre” and it holds in greater generality than stated here [1]. A direct proof when f is algebraic is as follows.

Proof.

$$\int_A^B p(f) dx = \sum_{\substack{\text{intervals } J \\ \text{where } f \text{ is} \\ \text{monotonic}}} \left(\int_J p(f) dx \right).$$

The sum is finite because f has a finite number of critical points. Now, substituting $y = (f|_J)(x)$,

$$\int_J p(f) dx = \int_{\text{image}(f|_J)} p(y) \frac{1}{|f'(f|_J^{-1}(y))|} dy.$$

Set

$$h_f^J(y) = \begin{cases} 0 & \text{if } y \notin f(J) \\ \frac{1}{|f'(f|_J^{-1}(y))|} & \text{if } y \in f(J). \end{cases}$$

Then we have

$$\int_A^B p(f) dx = \sum_J \int_{\text{image}(f)} p(y) h_f^J(y) dy = \int_{\text{image}(f)} p(y) \left(\sum_J h_f^J(y) \right) dy$$

and we see that

$$1_f(y) = \sum_J h_f^J(y)$$

satisfies

$$1_f(y) = \sum_{x \in f^{-1}(y)} \frac{1}{|f'(x)|}.$$

This proves the lemma.

Setting $p(y) =$ characteristic function of $(-\infty, y_0)$:

$$(1.1) \quad \text{measure} \{x | f(x) \leq y_0\} = \int_{f(x) \leq y_0} dx = \int_{-\infty}^{y_0} 1_f(y) dy.$$

Thus (a) $1_f \in L^1(\text{image}(f))$, and (b) $d\omega_f/dy = 1_f$. Also, from the definition of 1_f we see that (c) 1_f is analytic between poles which are the critical values of f . Now, (b) tells us that determining 1_f is equivalent to determining ω_f , so our problem is equivalent to determining f given 1_f .

2. The Main Theorem. For a simple statement of the main theorem we use a non-standard definition of critical value:

Define: The *critical values* of f are either

(1) The $f(x)$ where $f'(x) = 0$

or

(2) $f(A)$ and $f(B)$, when f is defined on $[A, B]$.

We say a critical value y_0 is *repeated* if there exists $x_1 \neq x_2$ such that $f(x_1) = f(x_2) = y_0$ and either $f'(x_i) = 0$ or $x_i \in \{A, B\}$ ($i = 1, 2$) (i.e. y_0 is a critical value in more than one way). If the critical values are *not* repeated we say they are *distinct*. A critical value y_0 is called *degenerate* if there is an x such that $f(x) = y_0$, $f'(x) = 0$, $f''(x) = 0$; if there is no such x , y_0 is called a *nondegenerate* critical value.

THEOREM. *If, on $[A, B]$, f is an analytic function with distinct non-degenerate critical values, then 1_f determines f .*

REMARK. To “determine f ” means that, up to translating and reflecting the domain, f is the unique analytic function defined on a finite interval, with distinct non-degenerate critical values, having this 1_f . Throughout below work with a fixed but unknown such analytic function f .

Proof. Setting the upper limit $y_0 = \infty$ in (1.1), we have

$$\int_{-\infty}^{\infty} 1_f dy = \text{measure}(\text{domain}(f)),$$

so the length of the interval of definition of f is known; so we may take $[A, B]$ to be any interval of this length.

Case 1. Assume $f'(A) \neq 0$ or $f'(B) \neq 0$. [A short proof of this case can be given; we give a slightly longer proof here in order to develop concepts used later.] Reflecting the domain through its midpoint if necessary, we may assume $f'(A) \neq 0$. *Claim:* in this case $f(A)$ is a non-pole discontinuity of 1_f .

Proof of Claim. Let J_1, J_2, \dots, J_k be the consecutive intervals on which f is monotonic (so $\inf(J_1) = A$, $\sup(J_k) = B$) and let

$$h_{J_i} = 1_{f|_{J_i}}.$$

Then

$$1_f = h_{J_1} + h_{J_2} + \dots + h_{J_k}.$$

Because the critical values are finite in number and distinct there is a neighborhood U of $f(A)$ containing no other critical values. Thus h_{J_i} is analytic in U for $i = 2, 3, \dots, k$. Now, h_{J_1} has a non-pole discontinuity at $f(A)$ because: if $f'(A) > 0$ then

$$\lim_{y \rightarrow f(A)^+} h_{J_1}(y) = \frac{1}{f'(A)}, \quad \lim_{y \rightarrow f(A)^-} h_{J_1}(y) = 0.$$

Similarly if $f'(A) < 0$, in which case

$$\lim_{y \rightarrow f(A)^-} h_{J_1}(y) = \frac{-1}{f'(A)}, \quad \lim_{y \rightarrow f(A)^+} h_{J_1}(y) = 0.$$

Thus since h_{J_1} has a non-pole discontinuity at $f(A)$ and the other h_J 's are analytic near $f(A)$, it follows that 1_f has a non-pole discontinuity at $f(A)$. $f(B)$ is the other possible non-pole discontinuity of 1_f , so we can find $f(A)$ or $f(B)$.

For simplicity assume $f'(A) > 0$, then we have for sufficiently small $\varepsilon > 0$ that

$$1_f = h_{J_2} + h_{J_3} + \cdots + h_{J_k} \quad \text{on } (f(A) - \varepsilon, f(A))$$

and

$$1_f = h_{J_1} + h_{J_2} + \cdots + h_{J_k} \quad \text{on } (f(A), f(A) + \varepsilon)$$

let $G =$ analytic continuation of $1_f|_{(f(A)-\varepsilon, f(A))}$ into $[f(A), f(A) + \varepsilon)$, then we have

$$G = h_{J_2} + h_{J_3} + \cdots + h_{J_k} \quad \text{on } (f(A) - \varepsilon, f(A) + \varepsilon).$$

Thus

$$1_f|_{(f(A), f(A)+\varepsilon)} - G|_{(f(A), f(A)+\varepsilon)} = h_{J_1}|_{(f(A), f(A)+\varepsilon)}$$

Now $h_{J_1}|_{(f(A), f(A)+\varepsilon)}$ is known so we know f near A :

$$\frac{d}{dy} (f|_{J_1})^{-1} = h_{J_1} \quad (\text{near } f(A)), \text{ i.e.}$$

$$(f|_{J_1})^{-1} = f(A) + \int_{f(A)}^y h_{J_1} dy \quad (\text{near } f(A)).$$

Therefore

$$f|_{J_1} = \left(f(A) + \int_{f(A)}^y h_{J_1} dy \right)^{-1} \quad (\text{near } A).$$

Since f is analytic, knowing it near A , there is a unique analytic continuation to the rest of $[A, B]$.

The same argument holds if $f'(A) < 0$ or if $f'(B) \neq 0$: 1_f jumps up at $f(A)$ (or $f(B)$) and we may subtract the analytic continuation of "before the jump" from "after the jump" and the difference is exactly (a restriction of) h_J where J is the interval containing the endpoint A (B). This being the derivative of $(f|_J)^{-1}$ we can compute $f|_J$ and then extend it to the rest of the domain by analytic continuation.

Case 2. $f'(A) = 0, f'(B) = 0$. The argument is the same once we have singled out the poles $f(A)$ and $f(B)$ from the others; to do this requires some ideas:

If y_0 is a pole of 1_f , since we assume non-degenerate and distinct critical values, y_0 is the value of a unique local maximum or minimum. It follows that the pole is one-sided i.e., if y_0 corresponds to a local maximum then

$$\lim_{y \rightarrow y_0^-} 1_f(y) = \infty \quad \text{and} \quad \lim_{y \rightarrow y_0^+} 1_f(y) < \infty,$$

or, if y_0 is the value of a local minimum then

$$\lim_{y \rightarrow y_0^+} 1_f(y) = \infty \quad \text{and} \quad \lim_{y \rightarrow y_0^-} 1_f < \infty.$$

For such a one-sided pole y_0 , choose $\epsilon > 0$ so small that $(y_0 - \epsilon, y_0 + \epsilon)$ contains no other critical values of f .

Define.

$$\epsilon(y_0) = \begin{cases} (y_0 - \epsilon, y_0) & \text{if } y_0 \text{ corresponds to a minimum} \\ (y_0, y_0 + \epsilon) & \text{if } y_0 \text{ corresponds to a maximum} \end{cases}$$

and

$$\epsilon'(y_0) = \begin{cases} (y_0, y_0 + \epsilon) & \text{if } y_0 \text{ corresponds to a minimum} \\ (y_0 - \epsilon, y_0) & \text{if } y_0 \text{ corresponds to a maximum.} \end{cases}$$

Thus 1_f has a pole in $\epsilon'(y_0)$ and none in $\epsilon(y_0)$; also $1_{f|_{\epsilon(y_0)}}$ has an analytic continuation into $\epsilon'(y_0)$ which is exactly $1_{f|_V}$ where $V = [A, B] \setminus$ (a neighborhood of the critical point corresponding to y_0).

Define. The pole part of 1_f at y_0 is

$$\pi_{y_0}(y) = 1_{f|_{\epsilon'(y_0)}} - \left(\text{analytic continuation to } \epsilon'(y_0) \text{ of } \left(1_{f|_{\epsilon(y_0)}} \right) \right).$$

This is defined for $y \in \epsilon'(y_0)$, and π_{y_0} is the part of 1_f arising from x 's near the critical point x_0 such that $f(x_0) = y_0$ (there are two such x 's if $x_0 \notin \{A, B\}$); the rest of 1_f is just the contribution from x 's not near the critical point x_0 ; which contribution, being present on both sides of y_0 can be subtracted from 1_f .

Each pole part arising from an interior critical point of $[A, B]$ consists of two terms: if x_0 is the interior critical point and if near x_0 we have x_- and x_+ as the two x 's such that

$$f(x_{\pm}) = y$$

then

$$\pi_{y_0}(y) = \frac{1}{|f'(x_-)|} + \frac{1}{|f'(x_+)|}.$$

These two terms are the *branches* of the pole part π_{y_0} . If $y_0 = f(A)$ or $f(B)$, the pole part is also defined, but consists of just a single branch.

Define. A critical value y_1 is a *successor* of y_0 if the analytic continuation of π_{y_0} has a pole at y_1 (i.e. if the maximal domain of definition of π_{y_0} is the open interval between y_0 and y_1).

LEMMA 4. *If $f'(A) = 0$ and $f'(B) = 0$, every critical value $y_0 = f(x_0)$ has exactly one successor $f(x)$; between x_0 and x there are no other critical points.*

Proof. Consider the consecutive intervals on which f is monotonic: J_i and J_{i+1} , where $J_i = [x_-, x_0]$. Defining $\varepsilon'(y_0)$ as before, we have for $y \in \varepsilon'(y_0)$ such that

$$\pi_{y_0}(y) = \frac{1}{|f'(f|_{J_i}^{-1}(y))|} + \frac{1}{|f'(f|_{J_{i+1}}^{-1}(y))|}.$$

We can analytically continue π_{y_0} by analytically continuing each of these two branches. The first is extended by $1_{f|_{J_i}}$ and the second by $1_{f|_{J_{i+1}}}$; thus the first extends to $f(x_-)$ and the second to $f(x_+)$. Since $f(x_-) \neq f(x_+)$ one is closer to $y_0 = f(x_0)$ and π_{y_0} extends analytically from y_0 to this nearer critical value. This proves the lemma: $f(x_0)$ has the successor $f(x_-)$ or $f(x_+)$; and there are no other critical points between x_0 and $x = x_{\pm}$.

One would hope that most critical values are successors and have a successor, and that $f(A)$ and $f(B)$ could be distinguished from the other poles as being the only critical values without both. Unfortunately, in some cases, a critical value is itself the successor of its successor; we still need more definitions.

Define A *block of poles* is an ordered sequence of poles of 1_f : y_0, y_1, \dots, y_s such that $y_i \neq y_j$ for $i \neq j$, such that for all i either y_i is the successor of y_{i+1} or y_{i+1} is the successor of y_i , and such that the sequence is *maximal* in the sense that it cannot be properly contained in any longer such sequence.

Our task is to discover which blocks of poles contain $f(A)$ and $f(B)$, and to single out those poles from the others of their respective blocks.

PROPOSITION 5. *If π_{y_i} and $\pi_{y_{i+1}}$ are the pole parts of successors then they may be analytically continued so as to share a branch.*

Proof. In the notation of above

$$\pi_{y_i} = \frac{1}{|f'(f|_{J_i}^{-1})|} + \frac{1}{|f'(f|_{J_{i+1}}^{-1})|}$$

$$\pi_{y_{i+1}} = \frac{1}{|f'(f|_{J_{i+1}}^{-1})|} + \frac{1}{|f'(f|_{J_{i+2}}^{-1})|}.$$

The second term of π_{y_i} is the same as the first term of $\pi_{y_{i+1}}$, but defined in different domains: $\varepsilon'(y_i)$ and $\varepsilon'(y_{i+1})$. Since y_i and y_{i+1} are in the relation of successors the maximal domain of extension of one pole part contains the maximal domain of extension of its successor. In the latter domain *both* pole parts can be defined and, there, they share the branch

$$\frac{1}{|f'(f|_{J_{i+1}}^{-1})|}.$$

In this domain into which both pole parts can be extended the difference

$$\pi_{y_i} - \pi_{y_{i+1}}$$

is defined and consists of the remaining branches of π_{y_i} and $\pi_{y_{i+1}}$, the shared branch cancels.

Suppose, for example, that y_i is an endpoint and y_{i+1} is its successor, then the difference $\pi_{y_i} - \pi_{y_{i+1}}$ is defined in the maximal domain of extension of $\pi_{y_{i+1}}$ or at least it can be extended to that domain and consists only of the *single* unshared branch of $\pi_{y_{i+1}}$. This remaining branch has a maximal domain of extension with a pole corresponding to the value of f at the third critical point counting in from the critical endpoint. If that pole is y_{i+2} then the maximal domain of extension of $\pi_{y_i} - \pi_{y_{i+1}}$ contains the maximal domain of extension of $\pi_{y_{i+2}}$; in this latter domain they share a branch and in this domain

$$(\pi_{y_i} - \pi_{y_{i+1}}) + \pi_{y_{i+2}}$$

is defined and, again, consists of one remaining branch of $\pi_{y_{i+2}}$.

If we could order the poles of 1_f according to the order of their corresponding critical points y_0, y_1, \dots, y_k . Then the alternating sum:

$$\pi_{y_0} - \pi_{y_1} + \pi_{y_2} \cdots + (-1)^k \pi_{y_k}$$

makes sense as follows: for each i

$$(\pi_{y_0} - \pi_{y_1} + \cdots + (-1)^i \pi_{y_i})$$

has a maximal domain of extension which contains the maximal domain of extension of $\pi_{y_{i+1}}$. In that domain they share a branch, and so the difference

$$\left(\pi_{y_0} - \pi_{y_1} \cdots + (-1)^i \pi_{y_i} \right) + (-1)^{i+1} \pi_{y_{i+1}}$$

is defined in this domain; again the shared branch cancels.

Digression. We can distinguish the poles corresponding the local maxima from those corresponding to local minima (recall the definition of $\varepsilon'(y_0)$) and can think of the sum as being

$$\sum_{\text{max's}} \pi_{y_i} - \sum_{\text{min's}} \pi_{y_j}.$$

It is an interesting observation that when we form this entire sum, *all* branches cancel and we are left with zero. The sum only makes sense however when the terms are added in order as before, in this order:

$$\sum_i (-1)^i \pi_{y_i} = 0.$$

This says no more than that the graph of the original function f is *connected*, it is a topological statement. Given a distribution ω_f and its derivative 1_f , we may ask: What are necessary and sufficient conditions that these arise from an analytic function f with distinct critical values? At this point we can see that the supposed 1_f must have the appropriate kinds of one-sided poles but, more subtly, it must also have this property: that the alternating sum (in some order) of its pole parts is zero. Since functions which are C^1 and analytic between critical points also have such pole-parts with the alternating sum identically zero, this condition is not sufficient to guarantee that f is globally analytic. However I believe it is sufficient if we consider such piecewise analytic functions. **End Digression**

Each block of poles (z_0, z_1, \dots, z_s) has two *external* poles z_0 and z_s . Amongst the external poles of all the blocks we must decide which correspond to an endpoint value $f(A)$ or $f(B)$. If y_i is an external pole of a block (or any pole for that matter) there exists a sequence of poles y_0, y_1, \dots, y_{i-1} such that the alternating sum

$$\pi_{y_0} - \pi_{y_1} + \cdots + (-1)^{i-1} \pi_{y_{i-1}}$$

is defined and has an extension with a pole at y_i and is “maximal” in the sense that the sequence y_0, y_1, \dots, y_{i-1} cannot be supplemented by including $y_{-k}, y_{-k+1}, \dots, y_{-1}$ s that

$$y_{-k}, \dots, y_{-1}, y_0, y_1, \dots, y_{i-1}$$

still has this property.

LEMMA 6. (a) *If y_i does not correspond to the value of an endpoint ($y_i \notin \{f(A), f(B)\}$) then y_0, y_1, \dots, y_{i-1} may be chosen outside of the block containing y_i .*

(b) *If $y_i \in \{f(A), f(B)\}$ then y_0, y_1, \dots, y_{i-1} must be chosen using some of the poles of the block containing y_i .*

Proof. If $y_i = f(A)$ or $f(B)$, it has a successor and is contained in some block containing more than just the critical value y_i . For a maximal sequence y_0, y_1, \dots, y_{i-1} such that

$$\sum (-1)^j \pi_{y_j}$$

is defined and has some domain extending only up to y_i , it is necessary that y_{i-1} be in the block of y_i . If y_i is not the value of an endpoint (A or B), but is the external pole of some block, we can choose y_0, y_1, \dots, y_{i-1} not contained in the block of y_i .

Now we are ready to finish the proof of the Theorem. Lemma 6 tells us how to find $f(A)$ and $f(B)$. Say $y_0 = f(A)$. Then setting

$$\tilde{\pi} = \text{analytic continuation of } \pi_{y_0}$$

we have

$$\tilde{\pi} = \frac{d}{dy} \left(f|_{[A, A+\epsilon]}^{-1} \right) \text{ near } y_0.$$

As before:

$$f|_{[A, A+\epsilon]}^{-1}(y) = f(A) + \int_{f(A)}^y \tilde{\pi}$$

so

$$f|_{[A, A+\epsilon]} = \left(f(A) + \int_{f(A)}^y \tilde{\pi} \right)^{-1}.$$

Thus, f is determined on $[A, B]$ by analytically continuing $f|_{[A, A+\epsilon]}$.

REFERENCES

[1] W. F. Pohl, *Some integral formulas for space curves and their generalizations*, Amer. J. Math., V. XC, No. 4 (Oct. 1968), p. 1328.

Received June 25, 1985.

UNIVERSITY OF SOUTHERN CALIFORNIA
 LOS ANGELES, CA 90089-1113

